

Parameter inference method for stochastic single-cell dynamics from tree-structured data

Irena Kuzmanovska, Andreas Miliadis-Argeitis, Christoph Zechner and Mustafa Khammash

July 6, 2016

Abstract

With the advance of experimental techniques such as time-lapse fluorescence microscopy, the availability of single-cell trajectory data has vastly increased, and so has the demand for computational methods suitable for parameter inference with this type of data. However, most of currently available methods treat single-cell trajectories independently, ignoring the mother-daughter relationships and the information provided by population structure. This information is however essential if a process of interest happens at cell division, or if it evolves slowly compared to the duration of the cell cycle. In this work, we highlight the importance of tracking cell lineage trees and propose a Bayesian framework for parameter inference on tree-structured data. Our method relies on a combination of Sequential Monte Carlo for likelihood approximation and Markov Chain Monte Carlo for parameter sampling. We demonstrate the capabilities of our inference framework on two simple examples in which the lineage tree information is necessary: one in which the cell phenotype can only switch at cell division and another where the cell type fluctuates randomly over timescales that extend well beyond the lifetime of a single cell.

1 Introduction

Biochemical processes in isogenic cells exhibit substantial heterogeneity [12, 5]. Understanding the latter demands for experimental techniques that can resolve such processes at the single-cell level. In contrast to bulk measurements, these techniques provide not only access to the average behavior of intracellular dynamics, but also its variability across cells and over time. Most single-cell techniques, however, reveal only very few components simultaneously that are often multiple steps away from the actual quantities of interest. The dynamics of a promoter, for instance, may not be accessible directly, but only indirectly through a fluorescent reporter that is expressed upon activation of this promoter [7]. Statistical inference in combination with mathematical models provides a means to reconstruct inaccessible parameters from available measurements, making it instrumental for studying biochemical processes based on single-cell data.

How such inference can be performed depends strongly on the way the data has been collected: flow cytometry measurements, for instance, reveal fluorescence values across a population but individual cells cannot be tracked over time. Consequently, measurements at two different time instances are considered statistically independent. Time-lapse microscopy techniques permit tracking of single cell trajectories over the duration of a whole experiment [17], which in turn provides a handle also on the temporal correlation of the underlying process. This additional degree of information can dramatically improve the inference of unknown process parameters [18].

Most existing inference approaches consider single cell trajectories to be statistically independent of each other [10, 18, 7, 2]. This way, however, important information stemming from the ancestry of a cell are lost: shortly after cell division, for example, two daughter cells are likely to exhibit substantial correlations, which cannot be captured by a model that assumes independence among cells. This can yield incomplete and biased results, especially when the time scale of the process under study is slow compared with the cell cycle duration.

In addition, stochastic processes of interest such as epigenetically regulated phase variation in bacteria are often driven by DNA replication just before cell division. Examples in this category are the regulation of *agn43* [16, 9] [9] and *Pap* [13, 14] systems in *E.coli*, and the glucosyltransferase (*gtr*) gene cluster in

Salmonella [3]. Due to the non-reversibility of the epigenetic modifications, gene replication (and consequently cell division) is crucial for switching to happen. Therefore, cell lineage information has to be taken into account in single-cell studies of these systems.

At present, there exist only little work on the statistical inference of tree-based single-cell data. In [11], the authors have proposed a method for inferring process parameters based on approximate Bayesian computation (ABC) from single-cell trajectories. While their approach applies in principle also tree-structured data, it requires all trajectories to have the same length and resolution. The method is therefore limited to cases where cell division times exhibit negligible variability across cells and over time.

In this work, we propose a Bayesian parameter inference framework for inferring parameters from general tree-structured data. Our method works also when the mother and daughter cells have different lifespan and different measurement points throughout their the window of acquisition. In Section 2 we will give a mathematical description of the inference problem and the class of models we consider, in Section 3 we will present our method in details and in Section 4 we will demonstrate our method on two different model examples by inferring some of their parameters.

2 Problem description

2.1 The model

To introduce the inference problem and the class of models considered here, we refer to the illustration in Figure 1. Let us consider an intracellular biochemical process of interest modeled by a continuous-time dynamical S . The system behavior within each cell can be monitored with the help of a dynamic readout, such as the abundance of a fluorescent reporter protein. Through time-lapse microscopy, we assume that a growing population of single-cells and their progeny can be tracked over time and measured at multiple time points (green dots in Figure 1), giving rise to a hierarchical tree data structure that describes the time evolution of the population.

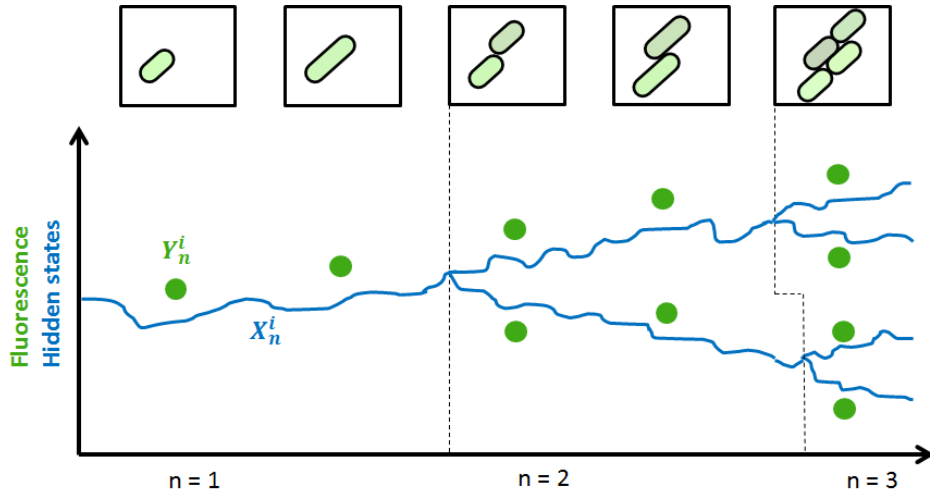


Figure 1: Graphical illustration of time-lapse microscopy images of a growing *E.coli* colony which contains a fluorescent reporter gene. After the cells are segmented and tracked, the fluorescence intensity in each of them can be extracted, giving rise to a fluorescence tree-structured dataset (greed dots). The continuous blue curve represents the unobserved state trajectory of each individual cell.

We assume that each tree starts with a single mother at generation 0 and that the population is followed over N generations, giving rise to a total number of 2^N cells at the final generation. The system S describes the evolution of a set of internal states x (blue curve in Figure 1). These states can be accessed partially at discrete time points through experimental techniques yielding a corresponding readout y . Each cell is

assigned a separate time index and a separate time of division, T , which can either be assumed known from the single-cell tracking data, or be inferred based on this data. We denote by \mathbf{X} the whole trajectory $\{x(t), t \in [0, T]\}$ from the time of birth of a cell (at $t = 0$) until its division (at $t = T$). The dynamics of S may evolve on a continuous, discrete or hybrid space, and similarly be stochastic, deterministic, or involve components of both types. In any case, we assume that S depends on a set of constant parameters Θ , which are assumed to be the same across the population.

At generation n there are 2^n cells in the population. From this point on, we will distinguish each cell by its generation number, n , and an index i , that ranges from 1 to 2^n . The i^{th} cell of the n^{th} generation gives rise to two daughters, indexed by $2i - 1$ and $2i$, in generation $n + 1$. Henceforth, all quantities related to a certain cell in a given lineage will be indexed by these two numbers.

Following this notation, we denote by \mathbf{X}_n^i the state trajectory of the i^{th} cell in the n^{th} generation; that is,

$$\mathbf{X}_n^i := \{x_n^i(t), t \in [0, T_n^i]\}.$$

The state trajectories of the daughters originating from a mother cell with state trajectory \mathbf{X}_n^i will therefore be denoted by \mathbf{X}_{n+1}^{2i-1} and \mathbf{X}_{n+1}^{2i} respectively. The corresponding discrete set of measurements associated with \mathbf{X}_n^i is denoted by \mathbf{Y}_n^i . More specifically,

$$\mathbf{Y}_n^i := \{y_n^i(t_{n,k}^i), k = 1, \dots, K_n^i\}.$$

This notation reflects the fact that the i^{th} cell of the n^{th} generation is observed at a total number of K_n^i time points (each denoted by $t_{n,k}^i$) during its lifetime, and that the number and location of observation time points will in general be different for every cell.

We will further denote by $P(x_{n+1}^{2i-1}(0), x_{n+1}^{2i}(0) | x_n^i(T_n^i), \Theta)$ the distribution of the daughter initial conditions given the state of the mother just before division, and call this the *transition probability* from one generation to the next. It is reasonable to assume that, once their respective initial conditions are determined based on their mother cell, the two daughters evolve independently of each other. As defined above, the transition probability mechanism may itself contain unknown parameters that need to be estimated from the data.

2.2 The inference problem

Our goal is to infer the posterior distribution of Θ given (1) the set of measured cellular readouts over the whole lineage, (2) our prior knowledge about Θ encoded in a prior distribution $\pi(\Theta)$ and (3) a measurement noise model that describes the likelihood of observing $y_n^i(t)$ given $x_n^i(t)$ (possibly also depending explicitly on unknown parameters contained in Θ). The latter is given by the density $f(y_n^i(t) | x_n^i(t), \Theta)$. With this measurement model, and assuming that measurements at individual time points are independent from each other, the likelihood of the whole measurement set for a single cell can be defined as

$$P(\mathbf{Y}_n^i | \mathbf{X}_n^i, \Theta) = \prod_{k=1}^{K_n^i} f(y_n^i(t_{n,k}^i) | x_n^i(t_{n,k}^i), \Theta).$$

Setting

$$\mathbf{X}_{tree} := \{\mathbf{X}_n^i, i = 1, \dots, 2^n, n = 0, \dots, N\}$$

and

$$\mathbf{Y}_{tree} := \{\mathbf{Y}_n^i, i = 1, \dots, 2^n, n = 0, \dots, N\},$$

the joint distribution over states and measurements over a tree starting from a single individual can be written as

$$\begin{aligned} P(\mathbf{X}_{tree}, \mathbf{Y}_{tree} | \Theta) &= P(\mathbf{X}_0^1 | \pi_x) P(\mathbf{Y}_0^1 | \mathbf{X}_0^1, \Theta) \\ &\times \prod_{n=1}^N \left[\left(\prod_{i=1}^{2^{n-1}} P(\mathbf{X}_n^{2i-1}, \mathbf{X}_n^{2i} | \mathbf{X}_{n-1}^i, \Theta) \right) \left(\prod_{i=1}^{2^{n-1}} P(\mathbf{Y}_n^i | \mathbf{X}_n^i, \Theta) \right) \right], \end{aligned} \quad (1)$$

where π_x is the initial distribution of $x_0^1(0)$. The likelihood of the measured outputs given Θ can therefore be obtained by marginalization of (1) over all possible unobserved states:

$$P(\mathbf{Y}_{tree} | \Theta) = \int P(\mathbf{X}_{tree}, \mathbf{Y}_{tree} | \Theta) d\mathbf{X}_{tree}. \quad (2)$$

As can be seen from the above equations, an additional difficulty of our inference problem in comparison to inference based on independent cell trajectories, is the fact that the likelihood $P(\mathbf{X}_{tree}, \mathbf{Y}_{tree} | \Theta)$ does not factorize over the readouts of individual cells, since the tree structure of the population introduces dependencies among the observations coming from different generations. The dependencies are generated through the unobserved state dynamics, which must therefore be taken into account.

Moreover, due to the dependencies introduced by the tree structure of the population, the integral in (2) is analytically intractable already for very simple state dynamics and its numerical evaluation scales exponentially with the number of generations in the tree. To address these difficulties, we employ a sequential Monte Carlo (SMC) scheme as described below to approximate the marginal likelihood (2).

3 Methods

3.1 Recursive likelihood and state posterior propagation

The joint distribution over states and observations given by (1) can be recursively computed, for example by first iterating over generations and then over the individuals of each generation. However, the same cannot be immediately said for (2), where the marginalization complicates the calculation. Here we propose an iterative calculation of this likelihood that again proceeds sequentially through the tree generations and the daughter pairs of each generation. As it turns out, marginalizing over the states introduces dependencies between different daughter pairs of the same generation that result in slightly more complicated formulas. These dependencies also complicate the numerical approximation of the likelihood, but, as we will see at the end of the section, this computation can be sped up considerably by making a reasonable simplifying approximation.

Before we derive the exact formulas, we need some additional notation. Let

$$\mathbf{Y}_n^{1,\dots,i} := \{\mathbf{Y}_n^1, \dots, \mathbf{Y}_n^i\}$$

denote the whole dataset of generation n and

$$\mathbb{Y}_{0:n} := \{\mathbf{Y}_m^{1,\dots,2^m}, m = 0, \dots, n\}$$

the dataset of all generations up to generation n . Similarly,

$$\mathbf{X}_n^{1,\dots,i} := \{\mathbf{X}_n^1, \dots, \mathbf{X}_n^i\}$$

and

$$\mathbb{X}_{0:n} := \{\mathbf{X}_m^{1,\dots,2^m}, m = 0, \dots, n\}.$$

To arrive at the exact formula for the likelihood, we first break up the total likelihood over the generations as follows (the dependence on Θ is suppressed to simplify the notation):

$$P(\mathbf{Y}_{tree}) = P(\mathbf{Y}_0^1) \prod_{n=1}^N P(\mathbf{Y}_n^{1,\dots,2^n} | \mathbb{Y}_{0:n-1}).$$

Assume now that $P(\mathbb{Y}_{0:n})$ (i.e. the likelihood of the subtree consisting of the first n generations) is available, and so is $P(\mathbb{X}_{0:n} | \mathbb{Y}_{0:n})$ (the state posterior over the same subtree). Consider the first two individuals of generation $n+1$, with state trajectories \mathbf{X}_{n+1}^1 and \mathbf{X}_{n+1}^2 , descending from the mother cell with state trajectory \mathbf{X}_n^1 .

Adding the information of this daughter pair to the posterior of the previous generations, we get

$$\begin{aligned}
P(\mathbf{X}_{n+1}^1, \mathbf{X}_{n+1}^2, \mathbb{X}_{0:n} | \mathbf{Y}_{n+1}^1, \mathbf{Y}_{n+1}^2, \mathbb{Y}_{0:n}) &= \frac{P(\mathbf{Y}_{n+1}^1, \mathbf{Y}_{n+1}^2, \mathbb{Y}_{0:n} | \mathbf{X}_{n+1}^1, \mathbf{X}_{n+1}^2, \mathbb{X}_{0:n}) P(\mathbf{X}_{n+1}^1, \mathbf{X}_{n+1}^2, \mathbb{X}_{0:n})}{P(\mathbf{Y}_{n+1}^1, \mathbf{Y}_{n+1}^2, \mathbb{Y}_{0:n})} = \\
&= \frac{P(\mathbf{Y}_{n+1}^1 | \mathbf{X}_{n+1}^1) P(\mathbf{Y}_{n+1}^2 | \mathbf{X}_{n+1}^2) P(\mathbf{X}_{n+1}^1, \mathbf{X}_{n+1}^2 | \mathbb{X}_{0:n})}{P(\mathbf{Y}_{n+1}^1, \mathbf{Y}_{n+1}^2 | \mathbb{Y}_{0:n})} \frac{P(\mathbb{Y}_{0:n} | \mathbb{X}_{0:n}) P(\mathbb{X}_{0:n})}{P(\mathbb{Y}_{0:n})} = \\
&= \frac{P(\mathbf{Y}_{n+1}^1 | \mathbf{X}_{n+1}^1) P(\mathbf{Y}_{n+1}^2 | \mathbf{X}_{n+1}^2) P(\mathbf{X}_{n+1}^1, \mathbf{X}_{n+1}^2 | \mathbf{X}_n^1)}{P(\mathbf{Y}_{n+1}^1, \mathbf{Y}_{n+1}^2 | \mathbb{Y}_{0:n})} P(\mathbb{X}_{0:n} | \mathbb{Y}_{0:n}).
\end{aligned}$$

The normalizing constant of the above posterior extends $P(\mathbb{Y}_{0:n})$ with the daughter pair of the next generation:

$$P(\mathbf{Y}_{n+1}^1, \mathbf{Y}_{n+1}^2 | \mathbb{Y}_{0:n}) = \iint \left(P(\mathbf{Y}_{n+1}^1, \mathbf{Y}_{n+1}^2 | \mathbf{X}_{n+1}^1, \mathbf{X}_{n+1}^2) P(\mathbf{X}_{n+1}^1, \mathbf{X}_{n+1}^2 | \mathbf{X}_n^1) d\mathbf{X}_{n+1}^1 d\mathbf{X}_{n+1}^2 \right) P(\mathbf{X}_n^1 | \mathbb{Y}_{0:n}) d\mathbf{X}_n^1$$

The above formulas allow us to update the starting posterior and likelihood with the first daughter pair from generation $n+1$. However, to add the second daughter pair (cells 3 and 4 of generation $n+1$, descending from cell 2 of generation n), we need to take into account the information provided by the first pair:

$$P(\mathbf{X}_{n+1}^{3,4}, \mathbf{X}_{n+1}^{1,2}, \mathbb{X}_{0:n} | \mathbf{Y}_{n+1}^{3,4}, \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n}) = P(\mathbf{X}_{n+1}^{1,2}, \mathbb{X}_{0:n} | \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n}) \frac{P(\mathbf{Y}_{n+1}^{3,4} | \mathbf{X}_{n+1}^{3,4}) P(\mathbf{X}_{n+1}^{3,4} | \mathbf{X}_{n+1}^{1,2}, \mathbb{X}_{0:n})}{P(\mathbf{Y}_{n+1}^{3,4} | \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n})}. \quad (3)$$

The above expression can be simplified by noting that $P(\mathbf{X}_{n+1}^{3,4} | \mathbf{X}_{n+1}^{1,2}, \mathbb{X}_{0:n}) = P(\mathbf{X}_{n+1}^{3,4} | \mathbb{X}_{0:n})$, i.e. daughter pairs of the same generation are conditionally independent given the parent states. However, the presence of the term $P(\mathbf{X}_{n+1}^{1,2}, \mathbb{X}_{0:n} | \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n})$ implies that, by taking into account the measurements of the first daughter pair, our posterior belief about the n -th generation states also needs to be updated before proceeding to the next pair. This leads to the creation of dependencies between the tree branches and means that they cannot be treated independently of each other, a feature that can create computational difficulties when one attempts to approximate the joint posterior by simulation. We thus make the *simplifying assumption* that

$$P(\mathbf{X}_n^i | \mathbf{Y}_{n+1}^{2i-1, 2i}, \mathbb{Y}_{0:n}) \approx P(\mathbf{X}_n^i | \mathbb{Y}_{0:n}).$$

In words, we assume that *the additional state information transferred from the measurement of a daughter pair at generation $n+1$ to their corresponding mother at generation n is negligible in comparison to the information provided by the previous generations to the mother*. As we will see, this allows us to treat each mother-daughter pair within a generation independently from the rest.

Continuing the analysis of the first two daughter pairs from above, we have that

$$\begin{aligned}
P(\mathbf{X}_{n+1}^{1,2}, \mathbb{X}_{0:n} | \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n}) &\approx P(\mathbf{X}_{n+1}^{1,2} | \mathbb{X}_{0:n}, \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n}) P(\mathbb{X}_{0:n} | \mathbb{Y}_{0:n}) = \\
&= P(\mathbf{X}_{n+1}^{1,2} | \mathbb{X}_{0:n}, \mathbf{Y}_{n+1}^{1,2}) P(\mathbb{X}_{0:n} | \mathbb{Y}_{0:n}).
\end{aligned}$$

This fact therefore leads to a simplification of the conditional likelihood, $P(\mathbf{Y}_{n+1}^{3,4} | \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n})$:

$$\begin{aligned}
P(\mathbf{Y}_{n+1}^{3,4} | \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n}) &= \\
&= \iint \left(P(\mathbf{Y}_{n+1}^{3,4} | \mathbf{X}_{n+1}^{3,4}) P(\mathbf{X}_{n+1}^{3,4} | \mathbf{X}_n^2) d\mathbf{X}_{n+1}^{3,4} \right) P(\mathbf{X}_{n+1}^{1,2} | \mathbf{X}_n^2, \mathbf{Y}_{n+1}^{1,2}) P(\mathbf{X}_n^2 | \mathbb{Y}_{0:n}) d\mathbf{X}_{n+1}^{1,2} d\mathbf{X}_n^2 = \\
&= \iint \left(P(\mathbf{Y}_{n+1}^{3,4} | \mathbf{X}_{n+1}^{3,4}) P(\mathbf{X}_{n+1}^{3,4} | \mathbf{X}_n^2) d\mathbf{X}_{n+1}^{3,4} \right) P(\mathbf{X}_n^2 | \mathbb{Y}_{0:n}) d\mathbf{X}_n^2 = P(\mathbf{Y}_{n+1}^{3,4} | \mathbb{Y}_{0:n}),
\end{aligned}$$

and the total likelihood of generation $n+1$ (conditioned on $\mathbb{Y}_{0:n}$) can be decomposed as a product of likelihoods over the individual daughter pairs.

Finally, the joint posterior (3) can be also decomposed as:

$$P\left(\mathbf{X}_{n+1}^{3,4}, \mathbf{X}_{n+1}^{1,2}, \mathbb{X}_{0:n} | \mathbf{Y}_{n+1}^{3,4}, \mathbf{Y}_{n+1}^{1,2}, \mathbb{Y}_{0:n}\right) = P\left(\mathbf{X}_{n+1}^{1,2} | \mathbf{Y}_{n+1}^{1,2}, \mathbb{X}_{0:n}\right) P\left(\mathbf{X}_{n+1}^{3,4} | \mathbf{Y}_{n+1}^{3,4}, \mathbb{X}_{0:n}\right) P\left(\mathbb{X}_{0:n} | \mathbb{Y}_{0:n}\right). \quad (4)$$

These facts will be put in use in the next section, where a Sequential Monte Carlo algorithm for the approximation of the tree likelihood will be presented.

3.2 Recursive likelihood approximation

Our SMC scheme is used to approximate $P(\mathbf{Y}_{tree} | \Theta)$, i.e. the likelihood of a set of measurements over a tree starting from a single individual, given a set of parameters Θ , under the simplifying assumption presented above. Our algorithm uses this assumption to exploit the conditional independence structure of the tree dynamics it generates in order to break down the likelihood computation. More concretely, the idea is to start at the root of the tree (i.e., a single cell) and recursively propagate the data likelihood from one generation to the next, treating each mother cell and its progeny independently. This can be understood as a generalization of recursive filtering for tree-structured data. To illustrate this better, we present the treatment of a single mother-daughter triplet in detail.

Assume that L samples (particles) from the known (prior) distribution of $x_n^i(T_n^i)$ are available. First, a pair of daughter cells is generated according to the transition probabilities $P(x_{n+1}^{2i-1}(0), x_{n+1}^{2i}(0) | x_n^i(T_n^i), \Theta)$ for each particle. Given the daughters' initial conditions, we next simulate each daughter until its own division time and calculate the likelihoods $P(\mathbf{Y}_{n+1}^{2i-1} | \mathbf{X}_{n+1}^{2i-1,l}, \Theta)$ and $P(\mathbf{Y}_{n+1}^{2i} | \mathbf{X}_{n+1}^{2i,l}, \Theta)$ for $l = 1, \dots, L$.

By assigning to the l -th particle a weight

$$w_{n+1}^{i,l} = P(\mathbf{Y}_{n+1}^{2i-1} | \mathbf{X}_{n+1}^{2i-1,l}, \Theta) P(\mathbf{Y}_{n+1}^{2i} | \mathbf{X}_{n+1}^{2i,l}, \Theta),$$

we next compute the marginal likelihood of the i^{th} daughter pair of generation $n+1$ by averaging the weights for all the particles:

$$P(\mathbf{Y}_{n+1}^{2i-1}, \mathbf{Y}_{n+1}^{2i} | \Theta) = \frac{1}{L} \sum_{l=1}^L w_{n+1}^{i,l}.$$

After normalizing the particle weights to sum up to one, we have thus obtained weighted samples from the posteriors $P(\mathbf{X}_{n+1}^{2i-1} | \mathbf{Y}_{n+1}^{2i-1}, \Theta)$ and $P(\mathbf{X}_{n+1}^{2i} | \mathbf{Y}_{n+1}^{2i}, \Theta)$. The samples are subsequently unweighted by resampling L particles from each posterior according to the normalized weights. These samples will serve as starting points for the daughters of the next generation. The same process is repeated for the rest of the n^{th} generation mothers, before moving on to generation $n+1$. This very general procedure is summarized in Algorithm 1.

Result: Estimate of $P(\mathbf{Y}_{tree}|\Theta)$

```

1 Create  $L$  replicates of  $\mathbf{X}_0^1, \{\mathbf{X}_0^{1,l}\}_{l=1}^L$  ;
2 Set  $P(\mathbf{Y}_{tree}|\Theta) = 1$  ;
3 for  $n = 0$  to  $N-1$  do
4   for  $i=1$  to  $2^n$  do
5     for  $l = 1$  to  $L$  do
6       Simulate a pair of daughter cells,  $\mathbf{X}_{n+1}^{2i-1,l}$  and  $\mathbf{X}_{n+1}^{2i,l}$  with initial conditions drawn from
         $P(x_{n+1}^{2i-1}(0), x_{n+1}^{2i}(0)|x_n^i(T_n^i), \Theta)$  ;
7       Compute the weights  $w_{n+1}^{i,l} = P(\mathbf{Y}_{n+1}^{2i-1,l} | \mathbf{X}_{n+1}^{2i-1,i}, \Theta)P(\mathbf{Y}_{n+1}^{2i,l} | \mathbf{X}_{n+1}^{2i,l}, \Theta)$  ;
8     end
9     Compute marginal likelihood of the triplet:  $P(\mathbf{Y}_{n+1}^{2i-1}, \mathbf{Y}_{n+1}^{2i}|\Theta) = L^{-1} \sum_{l=1}^L w_{n+1}^{i,l}$  ;
10    Update tree likelihood estimate:  $P(\mathbf{Y}_{tree}|\Theta) = P(\mathbf{Y}_{tree}|\Theta)P(\mathbf{Y}_{n+1}^{2i-1}, \mathbf{Y}_{n+1}^{2i}|\Theta)$  ;
11    Compute normalized weights:  $\tilde{w}_{n+1}^{i,l} := w_{n+1}^{i,l} / \sum_{l=1}^L w_{n+1}^{i,l}$  ;
12    Resample  $\{\mathbf{X}_{n+1}^{2i-1,l}\}_{l=1}^L$  and  $\{\mathbf{X}_{n+1}^{2i,l}\}_{l=1}^L$  according to  $\{\tilde{w}_{n+1}^{i,l}\}_{l=1}^L$  ;
13  end
14 end

```

Algorithm 1: The SMC algorithm for tree likelihood calculation

Notes:

1. Since we assume that the measurements from individual trees are independent from each other, the joint likelihood of a dataset consisting of several trees is simply a product of the likelihood of the individual trees. The likelihoods of individual trees can be thus estimated in parallel. Moreover, looking at the algorithm structure for a single tree, the likelihood calculation can be parallelized at two levels: 1) the mother cells of a given generation can be treated independently of each other 2) individual particle calculations for a given mother-daughter triplet can be done in parallel.
2. When no randomness is present in the transition mechanism from the mother to the daughters, the daughter cells can be treated completely independently given the state of the mother. This implies that they can be assigned independent weights at line 7 of Algorithm 1 : $w_{n+1}^{2i-1,l} = P(\mathbf{Y}_{n+1}^{2i-1,l} | \mathbf{X}_{n+1}^{2i-1,i}, \Theta)$ and $w_{n+1}^{2i,l} = P(\mathbf{Y}_{n+1}^{2i,l} | \mathbf{X}_{n+1}^{2i,i}, \Theta)$. In this way, the marginal likelihood computation of line 9 can be written as $P(\mathbf{Y}_{n+1}^{2i-1}, \mathbf{Y}_{n+1}^{2i}|\Theta) = \left(L^{-1} \sum_{l=1}^L w_{n+1}^{2i-1,l}\right) \left(L^{-1} \sum_{l=1}^L w_{n+1}^{2i,l}\right)$. Additionally, the normalized weights for the particle populations of the two daughters (line 11) are now independent: $\tilde{w}_{n+1}^{2i-1,l} := w_{n+1}^{2i-1,l} / \sum_{l=1}^L w_{n+1}^{2i-1,l}$ and $\tilde{w}_{n+1}^{2i,l} := w_{n+1}^{2i,l} / \sum_{l=1}^L w_{n+1}^{2i,l}$, which finally implies that the populations $\{\mathbf{X}_{n+1}^{2i-1,l}\}_{l=1}^L$ and $\{\mathbf{X}_{n+1}^{2i,l}\}_{l=1}^L$ (line 12) can be resampled independently.
3. The most computationally intensive step of the algorithm lies between lines 5-9, where the marginal likelihood of each mother-daughter triplet needs to be computed. Depending on the type of the unobserved state dynamics, accurate marginalization may require the use of very large particle numbers and greatly increase the computational cost of the algorithm. Typically, the situation is worse when the hidden state contains components driven by stochastic dynamics. This challenge has already been recognized and addressed in the literature, since it also appears in the parameter inference problem from independent single-cell trajectories [2, 1]. One can thus employ one of the several available alternatives at this step, such as sequential computation of the likelihood [2], or the use of approximating dynamics [2, 15, 6].

3.3 A pseudo-marginal MCMC sampler for parameter inference

Let us denote with $\bar{\Theta}$ the set of all unknown parameters from the vector Θ that we would like to infer. The goal of Bayesian inference is to compute or approximate via sampling the posterior distribution of these

parameters, $P(\bar{\Theta}|\mathbf{Y}_{tree}) \propto P(\mathbf{Y}_{tree}|\bar{\Theta})\pi(\bar{\Theta})$. To this end, we follow the “pseudo-marginal” MCMC approach [4], according to which a Markov Chain Monte Carlo (MCMC) sampler makes use of the noisy marginal likelihood estimates provided by the SMC algorithm of Section 3.2 to generate samples from the posterior of $\bar{\Theta}$. Crucially, despite the presence of noise in the SMC-based likelihood estimation, the MCMC sampler is still able to target the correct distribution. This property is in fact guaranteed for any marginal likelihood estimator that provides unbiased estimates of the likelihood [4], although the mixing properties of the MCMC chain will of course depend on the estimate variability. Given the fact that the SMC algorithm employed in this work indeed provides unbiased marginal likelihood estimates [8], the pseudo-marginal MCMC scheme described in Algorithm 2 targets the correct posterior parameter distribution. In addition to the pseudo code presented, we summarize our inference framework in Figure 2.

Result: $\{\bar{\Theta}_m\}_{m=1}^M \sim P(\bar{\Theta}|\mathbf{Y}_{tree})$

- 1 Draw an initial point $\bar{\Theta}_1$ from the prior $\pi(\bar{\Theta})$;
- 2 Estimate the likelihood $P(\mathbf{Y}_{tree}|\bar{\Theta}_1)$ using Algorithm 1;
- 3 **for** $m = 2$ **to** M **do**
- 4 Propose a parameter vector $\bar{\Theta}^*$ according to a proposal distribution $q(\bar{\Theta}^*|\bar{\Theta}_{m-1})$;
- 5 Calculate the likelihood $P(\mathbf{Y}_{tree}|\bar{\Theta}^*)$ using Algorithm 1 ;
- 6 Sample $u \sim \mathcal{U}([0, 1])$;
- 7 If

$$u < \min\left\{1, \frac{P(\mathbf{Y}_{tree}|\bar{\Theta}^*) \cdot q(\bar{\Theta}_{m-1}|\bar{\Theta}^*)}{P(\mathbf{Y}_{tree}|\bar{\Theta}_{m-1}) \cdot q(\bar{\Theta}^*|\bar{\Theta}_{m-1})}\right\},$$

accept the proposed parameters and set $\bar{\Theta}_m = \bar{\Theta}^*$ and $P(\mathbf{Y}_{tree}|\bar{\Theta}_m) = P(\mathbf{Y}_{tree}|\bar{\Theta}^*)$; else, set $\bar{\Theta}_m = \bar{\Theta}_{m-1}$ and $P(\mathbf{Y}_{tree}|\bar{\Theta}_m) = P(\mathbf{Y}_{tree}|\bar{\Theta}_{m-1})$.
- 8 **end**

Algorithm 2: The pseudo-marginal MCMC sampler for parameter inference

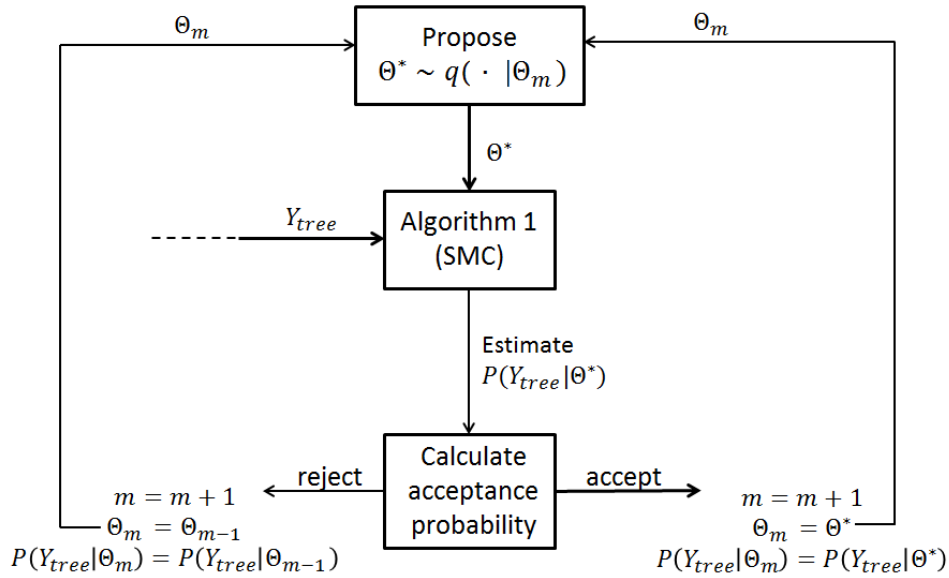


Figure 2: An overview of our inference framework

It is important to note that, in order for the above sampler to target the correct posterior despite the noisy likelihood estimates, it is necessary to keep the *previous* estimate of the likelihood in the acceptance ratio whenever a proposed parameter vector is rejected, and not get another SMC estimate for the same

point [4]. Fortunately, this also reduces the computational burden of the sampler, as it requires one likelihood estimation per iteration. On the negative side, the use of very noisy SMC estimates may considerably slow down the mixing of the sampler, since the chain may get trapped at a point with artificially large likelihood value. However, as the variance of the estimator decreases (e.g. through the use of larger particle sample sizes), it is expected that the mixing speed of our sampler will converge to that of a sampler with perfect (i.e. noiseless) marginal likelihood information.

4 Applications and results

In the following sections we will consider two possible examples for the dynamical system S and demonstrate the application of our inference method on these cases. In both examples, we assume that a cell is characterized by a discrete state, x_d . Over time and across generations, cells stochastically adopt a certain *type* determined by x_d . The cell type in turn determines the evolution of a continuous state vector, x_c , which may for example correspond to the immature and mature molecule types of a fluorescent reporter. In abstract terms, x_d may be thought of as the state a gene, whose activity affects the cell phenotype.

In the first example, the discrete state dynamics is described by a generalized two-type branching process. According to this scenario, the type of a cell is fixed throughout its lifetime and may change only at cell division, since the types of the daughters depend probabilistically on the type of the mother. In the second example, the cell type may stochastically vary throughout the cell lifetime according to a two-state continuous-time Markov chain, while the two daughters are assumed to inherit the type of the mother. To test the performance of our inference framework, we generated simulated datasets for the two example systems and used them to infer parameters of interest in each case. Details about some of the parameters used in the data generation process are provided in Table 1. The results for each example system are summarized below.

Table 1: Parameters used in the synthetic data generation

General	Example 1	Example 2
Number of trees	2	1
Number of generations	5	5
Cell division time (<i>min</i>)	30	30
Measurement interval (<i>min</i>)	5	5
Measurement model related		
GFP production rates for OFF type $\alpha_{OFF}(a.u.min^{-1})$	0.2	0.05
GFP production rates for ON type $\alpha_{ON}(a.u.min^{-1})$	1	20
GFP maturation rate $m(min^{-1})$	0.0462	0.0462
GFP dilution rate $d(min^{-1})$	0.0261	0.0231
GFP to fluorescence scaling constant c	100	100
measurement variance σ^2	500	500

4.1 Example 1: A two-type branching process with dynamic readouts

In this example, cells can adopt one of two possible types (ON or OFF) and maintain their type throughout their lifetime, which, for simplicity, we assume to be the same and equal to T for every cell. At cell division, the daughter cell types are determined based on the type of the mother cell, according to a set of transition probabilities, as illustrated in Figure 3A. In turn, the type of each cell is assumed to determine the production rate of a fluorescent reporter protein (such as GFP) which can then be observed using fluorescence time-lapse microscopy.

The state vector of each cell is thus defined as $x = [x_d \ G_{imm} \ G_{mat}]$, where x_d contains the cell type, while G_{imm} and G_{mat} correspond to the concentrations of the immature (dark) and mature (fluorescent) forms of the fluorescent reporter. Out of these, we assume that we can only obtain noisy measurements of G_{mat} at discrete points in time. Contrary to the cell type, the concentrations of the two reporter species are carried over from the mother to the daughters unchanged. That is, $G_{imm,n+1}^{2i-1}(0) = G_{imm,n}^i(T)$, $G_{mat,n+1}^{2i-1}(0) = G_{mat,n}^i(T)$ and similarly for the second daughter. This is a reasonable modeling assumption,

given that a daughter cell has half the volume of the mother and receives roughly half of its protein content as well.

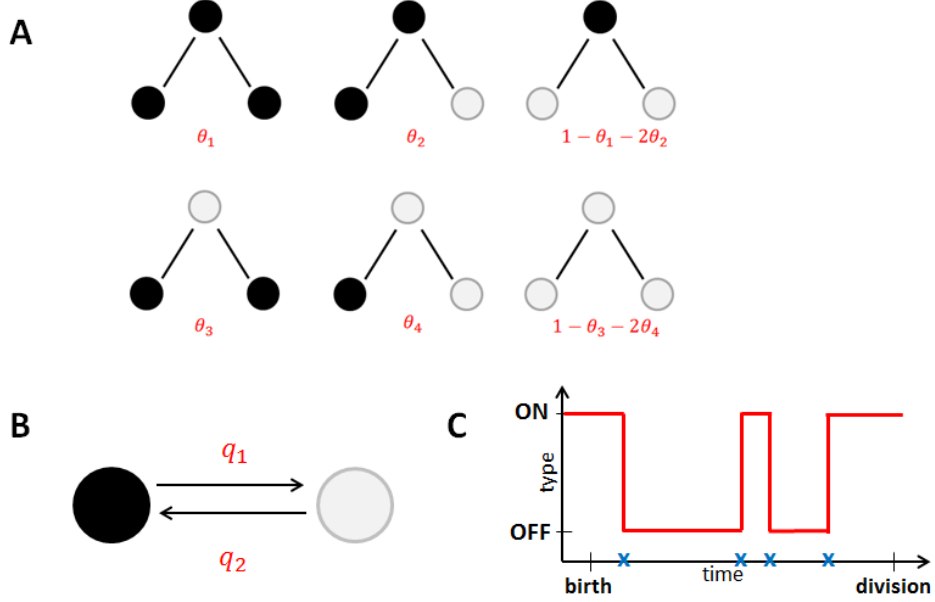


Figure 3: A.) **Example 1** All possible daughter pairs from a single mother and the corresponding probabilities of obtaining those pairs to be inferred. The different cell types here are black and white (OFF and ON). Note that the probability of obtaining first daughter cell black and the second white is same as obtaining the first daughter cell white and the second black. B.) **Example 2:** Cell types switch during the cell lifetime according to a two state CTMC with rates q_1 and q_2 , C.) An example of the evolution of a cell type during its lifetime. The blue crosses on the time axis indicate switching point from ON-OFF or vice versa.

The fluorescent reporter dynamics of every cell evolves according to the following set of linear ODEs:

$$\dot{G}_{imm} = \alpha(x_d) - \delta \cdot G_{imm} - m \cdot G_{imm} \quad (5)$$

$$\dot{G}_{mat} = -\delta \cdot G_{mat} + m \cdot G_{imm}, \quad (6)$$

where $\alpha(x_d)$, δ and m are reporter protein production, dilution and maturation rates respectively. The production rate is determined by the cell type: for an OFF-type cell, $\alpha(OFF) = \alpha_{OFF}$, while a cell of the ON type has $\alpha(ON) = \alpha_{ON} > \alpha_{OFF}$.

As described above, we assume that noise-corrupted measurements proportional to the G_{mat} species are available at M points, t_1, \dots, t_M , during the life of every cell. The readout of a single cell at a given time is therefore assumed to be a scaled and noisy version of the G_{mat} concentration:

$$y(t_m) \propto \mathcal{N}(c \cdot G_{mat}, \sigma^2 \cdot G_{mat}), \quad (7)$$

where c and the measurement variance σ^2 are known. Given that individual measurements for each cell are independent from each other, the expression for the likelihood $P(\mathbf{Y}_n^i | \mathbf{X}_n^i, \Theta)$, where $\mathbf{Y}_n^i = \{y_n^i(t_m), m = 1, \dots, M\}$, is given by

$$P(\mathbf{Y}_n^i | \mathbf{X}_n^i) = \prod_{m=1}^M P(y_n^i(t_m) | G_{mat,n}^i(t_m)). \quad (8)$$

Using this type of reporter measurements for every cell belonging to a fully observed tree spanning N generations, our goal is to infer: 1. The transition probabilities that govern cell type switching ($\theta_1, \dots, \theta_4$ in Fig. 3A) and 2. The type-specific reporter production rates α_{OFF} and α_{ON} .

If the type of each cell was readily measurable, the use of simple maximum likelihood estimators for branching processes would suffice to obtain all the necessary discrete state statistics from a fully observed tree, making the use of the reporter model unnecessary. However, the intervening reporter maturation step, the slow dilution dynamics and the sparse, noisy sampling, make inference much more challenging and require the use of the sophisticated computational machinery presented in this work.

To test the performance of our algorithm on this system we simulated dataset in which we generated the daughter types according to the transition probabilities shown in Table 2, with GFP reporter production rates $\alpha_{OFF} = 0.2 \text{ a.u.min}^{-1}$ and $\alpha_{ON} = 1 \text{ a.u.min}^{-1}$.

		Daughter Types			
		(OFF,OFF)	(OFF,ON)	(ON,OFF)	(ON,ON)
Mother type	OFF	$\theta_1 = 0.6$	$\theta_2 = 0.1$	θ_2	$1 - \theta_1 - 2\theta_2$
	ON	$\theta_3 = 0.1$	$\theta_4 = 0.05$	θ_4	$1 - \theta_3 - 2\theta_4$

Table 2: Transition probabilities for the cell types considered in Example 1 and depicted on Figure 3.

Note that due to symmetry, the second and third entries of each row are equal. Moreover, the values of the first and second entries in each row determine the rest of the entries, since every row sums to one. We therefore considered θ_1 , θ_2 , θ_3 and θ_4 as unknown, together with the reporter production rates at each state, $\theta_5 := \alpha_{OFF}$ and $\theta_6 := \alpha_{ON}$. For all unknown parameters, we considered flat priors supported on the appropriate sets ($[0, 1]$ in the case of transition probabilities, and $[0, +\infty]$ for the production rates). The rest of the system parameters were fixed at the values reported in Table 1.

We ran the pseudo-marginal MCMC sample (Algorithm 2) to generate samples from the posterior distribution of $\bar{\Theta} = [\theta_1 \ \theta_2 \ \theta_3 \ \theta_4 \ \alpha_{OFF} \ \alpha_{ON}]$, using appropriate proposal densities for the different types of parameters: transition probabilities θ_1 and θ_2 were sampled from the three-dimensional simplex with the help of a Dirichlet distribution with its mode located at the current parameter values. Similarly, θ_3 and θ_4 were sampled independently from a second Dirichlet distribution, whereas θ_5 and θ_6 were generated from independent log-normal distributions with log-means equal to $\log(\theta_5)$ and $\log(\theta_6)$ respectively, and log-standard deviation equal to 0.02. The estimated posterior distributions $P(\bar{\Theta}|Y)$ based on 1500 MCMC steps are given in Figure 6, where it can be clearly seen that the inferred posterior means (black dashed lines) are located close to the true parameter values (red lines).

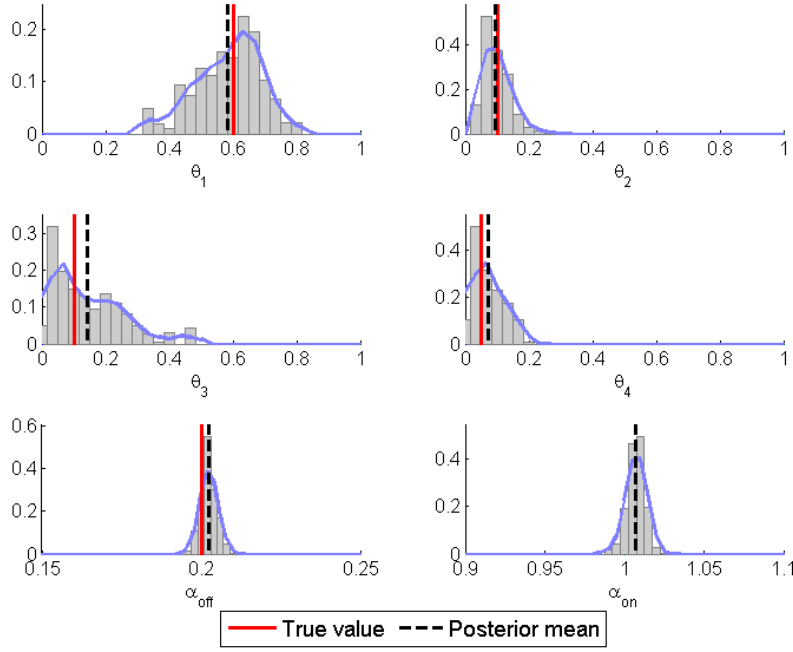


Figure 4: Histograms of sampled posterior distributions for the six unknown parameters (after the burn-in of the first 300 iterations had been discarded). The red vertical bar is positioned at the true parameter value (the one used for data generation), while the black dashed line is positioned at the estimated posterior mean (they are overlapping in the last plot). The blue curves are obtained by smoothing of the histograms.

It is known from theory that the SMC estimator for the likelihood approximation is unbiased [8]. To assess its variability as a function of the number of particles and thus determine approximately the particle number required for sufficiently accurate likelihood estimation in the MCMC sampler, we estimated the likelihood of a small tree $P(Y_{tree}|\bar{\Theta})$ with different numbers of particles given the true parameter values Θ_{true} . As can be seen on Figure 5, the average of 100 likelihood estimates fluctuates considerably for small particle number, but stabilizes as the particle number increases. With the increase of the number of particles the coefficient of variation of the likelihood calculations also drops quickly.

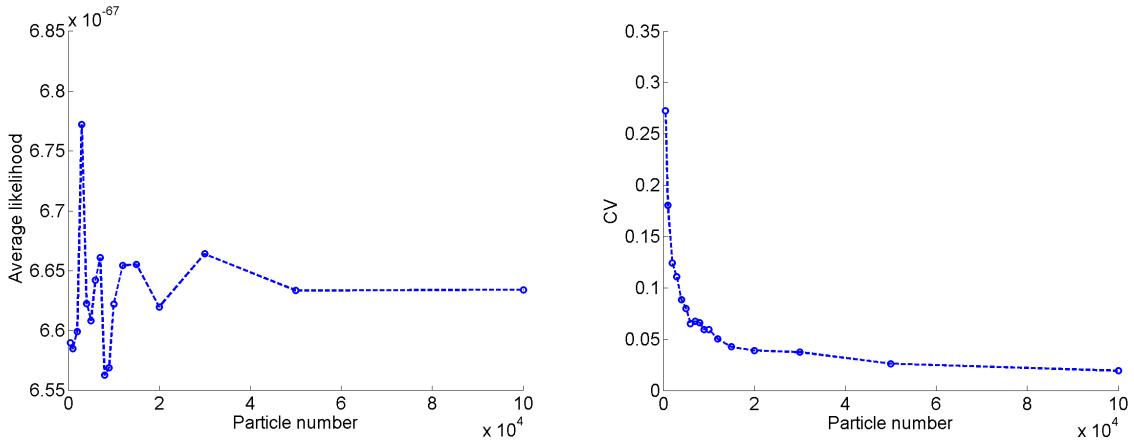


Figure 5: Mean of 100 estimates of $P(Y_{tree}|\bar{\Theta}_{true})$ with SMC vs. number of particles used (left) and their CV (right).

4.2 Example 2: Stochastic cell type switching

In the second example, we assume that the cell type evolves according to a two-state continuous-time Markov chain (CTMC) with rates q_1 and q_2 for the OFF-to-ON and ON-to-OFF transitions respectively. At division, each daughter inherits the type of its mother (together with the reporter concentrations, as before), but subsequently evolves independently from other cells according to the CTMC dynamics, as shown on Figure 3B and C. In this case, the reporter production rate alternates between $\alpha(ON)$ and $\alpha(OFF)$ in accordance with the cell type.

Using the same type of reporter model dynamics and readouts as in the previous example, our goal in this case it to infer the CTMC transition rates, q_1 and q_2 (Figure 3 B), while assuming the rest of the parameters known.

While in the first example system the tree-structure information was essential for inference as the cell types can only change at cell division, it is not equally obvious that our method will outperform traditional inference on independent single cell trajectories in this example system, where each daughter inherits the state of its mother and then evolves independently. To verify this, we additionally performed parameter inference by breaking up the tree into individual cell trajectories and considering each cell independently from the others, as it is usually done in conventional inference methods for single-cell data.

For this second type of inference, the priors for the reporter initial conditions of the daughters were determined according to the first measurement point of each cell, by discarding the part of its trajectory between the mother division and the first measurement. More specifically, the initial mature reporter (G_{mat}) value for each particle was extracted from the first available measurement point, by dividing the fluorescence with the scaling constant c . The immature reporter value (G_{imm}) was then drawn independently for each particle from a log-normal distribution with log-mean equal to $\log(G_{mat} \cdot \frac{d}{m})$ and appropriately tuned log-variance. With respect to the discrete states, a uniform prior on the two cell types was assumed.

For the MCMC sampler, we used two independent log-normal proposal kernels with log-means equal to the logarithm of the current parameter values at each step, with standard deviation of 0.2. Figure 6 compares the posterior distributions of the parameters for different datasets, obtained when the inference was done both on trees and on individual trajectories.

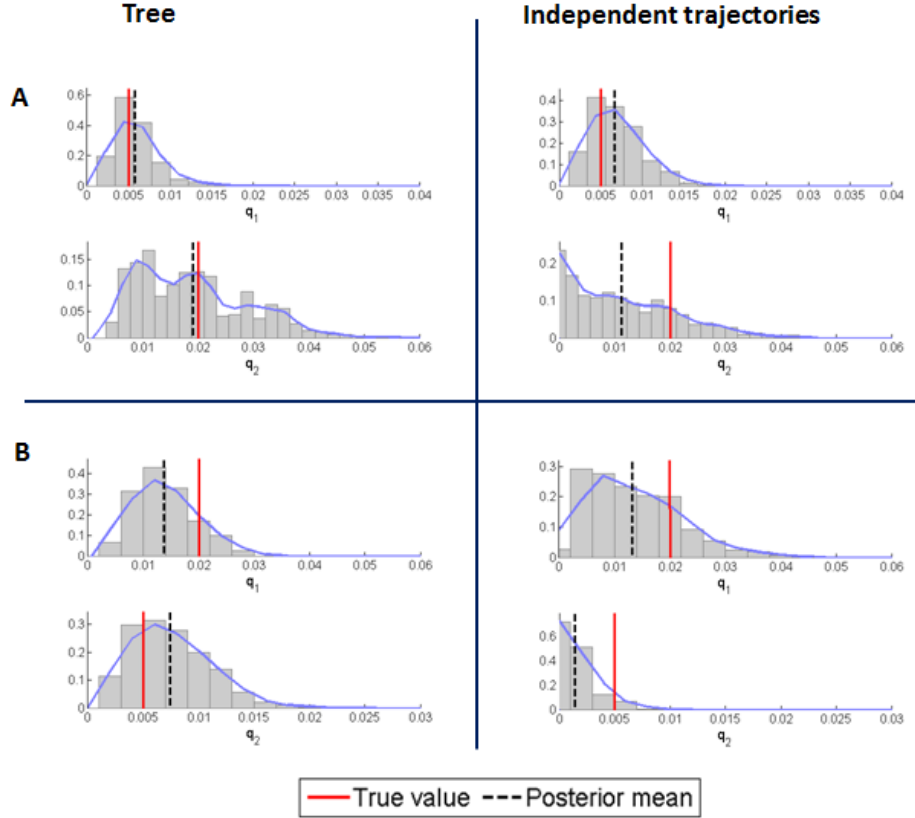


Figure 6: Histograms of sampled posterior distributions of the two unknown switching rates in Example 2 (q_1 : switching rate from OFF to ON and q_2 : switching rate from ON to OFF). The inference results based on tree-structured data (left column) are compared with the results obtained from independent cell trajectories (right column), as described in the text. Simulated data was generated using different parameter sets in panels A. and B. The red bars are positioned at the true parameter values (i.e. the ones used for data generation), while the black dashed lines indicate the estimated posterior mean. The blue curves are obtained by smoothing of the histograms.

One of the advantage of using tree-structured data is that the uncertainty regarding the initial conditions of each cell is greatly reduced, since the prior of each daughter state is based on the posterior of its mother. On the contrary, when the cell trajectories are assumed to be independent, the state of every cell has to be independently initialized according to the assumed prior. Moreover, when the switching rates are very small (corresponding to mean holding times that are close to or exceed the lifetime of a single cell), inference based on single-cell trajectories tends to underestimate them. This happens because information on the Markov chain state is only available during the lifetime of a cell. This is clearly not sufficient for precise estimation of the switching rates, which can only be upper-bounded by the data in this case.

In the case of tree-structured data on the other hand, the daughter cells inherit the state of their mothers, which implies that the history of a cell's lineage is taken into an account in the inference. Practically, this is equivalent to observing coupled sample paths of the Markov chain over longer time scales, which can clearly provide more information about the switching rates. On the contrary, when the switching rates are larger and correspond to mean holding times comparable to or shorter than the lifetime of a cell, inference on independent trajectories performs similarly to the inference on trees.

5 Discussion

In this work we proposed a parameter inference method for stochastic single-cell dynamics from tree-structured data. More specifically, we considered a class of systems with one or more unobserved states and fluorescent reporter readouts, observed through time-lapse microscopy, which allows tracking individual cells and their progeny over time. Our goal was to estimate the posterior distribution of the unknown system parameters given such readouts. To calculate the likelihood of the data for a given parameter set, the hidden state trajectories had to be integrated out. This marginalization was accomplished with the help of a Sequential Monte Carlo (SMC) method, which recursively computes a sampling-based estimate of this analytically intractable quantity. To sample the system parameter space, we employed an MCMC scheme, which is guaranteed to target the correct parameter posterior despite the noisy likelihood estimates. The application of our method to two simple examples showed that it can correctly infer the parameters of interest and approximate their posterior distribution. Our inference framework extends to more complex applications in a straightforward manner, although its computational complexity increases with the state and parameter dimensionality. In particular, more complex systems require a larger number of particles to achieve a reasonable accuracy of the SMC-based likelihood estimator. If the latter is too noisy, one may observe slow mixing of the MCMC sampler and in turn poor posterior estimates. Currently, our inference algorithm is implemented in Matlab and long runs with many particles are computationally very expensive. Encoding of the algorithm into more powerful programming languages and parallelization of the sampling process can dramatically increase its performance and enable its application to more complex models and larger datasets.

Our inference framework could be very useful in the case of systems where accurate tracking of single-cell dynamics across cell lineages plays an important role. These are, for example, systems involved in stem cell fate decisions, or stochastic phenotype switching in bacteria [16],[13]. In many such cases, stochastic fluctuations of key factors over long timescales and/or stochastic events taking place at cell division create strong mother-daughter and daughter-daughter correlations that play a crucial role in determining the overall behavior of a colony. In such cases, treatment of the measured single-cell trajectories independently from each other will result in a large loss of information and biased parameter estimates. We believe that proper incorporation of the population lineage information into the parameter inference problem will thus provide the right framework for treating this type of systems and may reveal important insights into their function.

References

- [1] Michael Amrein and Hans R. Künsch. Rate estimation in partially observed Markov jump processes with measurement errors. *Statistics and Computing*, 22(2):513–526, 2011.
- [2] Darren J. Wilkinson and Andrew Golightly. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*, 2011.
- [3] S. E. Broadbent, M. R. Davies, and M. W. van der Woude. Phase variation controls expression of salmonella lipopolysaccharide modification genes by a dna methylation-dependent mechanism. *Molecular Microbiology*, 77(2):337–353, 7 2010.
- [4] Gareth O. Roberts and Christophe Andrieu. The Pseudo-Marginal Approach for Efficient Monte Carlo Computations. *The Annals of Statistics*, 37(2):697–725, 2009.
- [5] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–6, August 2002.
- [6] A. Golightly and D. J. Wilkinson. Bayesian Inference for Stochastic Kinetic Models Using a Diffusion Approximation. *Biometrics*, 61(3):781–788, 2005.
- [7] Anders S Hansen and Erin K O’Shea. Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Molecular Systems Biology*, 9(1), 2013.
- [8] Hans R. Künsch. Particle filters. *Bernoulli*, 19(4):1391–1403, 09 2013.

- [9] Han N Lim and Alexander Van Oudenaarden. A multistep epigenetic switch enables the stable inheritance of dna methylation states. *Nature genetics*, 39(2):269–275, 2007.
- [10] Art  mis Llamosi, Andres Gonzalez-Vargas, Cristian Versari, Eugenio Cinquemani, Giancarlo Ferrari-Trecate, Pascal Hersen, and Gregory Batt. What population reveals about individual cell identity: Single-cell parameter estimation of models of gene expression in yeast. *PLoS Computational Biology*, 2015.
- [11] Carolin Loos, Carsten Marr, Fabian J Theis, and Jan Hasenauer. Approximate Bayesian Computation for Stochastic Single-Cell Time-Lapse Data Using Multivariate Test Statistics. In *Proceedings of the 13th International Conference on Computational Methods in Systems Biology*, pages 52–63, 2015.
- [12] Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA*, 94(3):814–819, 1997.
- [13] Brian Munsky, Aaron Hernday, David Low, and Mustafa Khammash. Stochastic modeling of the pap-pili epigenetic switch. *Proc. FOSBE*, pages 145–148, 2005.
- [14] Peter O’Hanley, David Low, Irevis Romero, David Lark, Kenneth Vosti, Stanley Falkow, and Gary Schoolnik. Gal-gal binding and hemolysin phenotypes and genotypes associated with uropathogenic escherichia coli. *New England Journal of Medicine*, 313(7):414–420, 1985.
- [15] Vassilios Stathopoulos and Mark A. Girolami. Markov chain Monte Carlo inference for Markov jump processes via the linear noise approximation. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371(1984), 2012.
- [16] Marjan W. van der Woude and Ian R. Henderson. Regulation and function of ag43 (flu). *Annual Review of Microbiology*, 2008.
- [17] Jonathan W Young, James CW Locke, Alphan Altinok, Nitzan Rosenfeld, Tigran Bacarian, Peter S Swain, Eric Mjolsness, and Michael B Elowitz. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature protocols*, 7(1):80–88, 2012.
- [18] Christoph Zechner, Michael Unger, Serge Pelet, Matthias Peter, and Heinz Koepl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods*, 11(2):197–202, 2014.